Bike Rental Predictive Modeling Project

Samantha Simpson

2021-12-10

Problem Setup

For this project, the dataset that I chose is the 'Bike Sharing Dataset,' found at https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset. The response variable for this dataset is the number of bike rentals per day, and the predictor variables are a set of seasonal and environmental factors that may affect the number of bike rentals.

The response variable for this dataset is the total number of bike rentals per day. Therefore, I felt that a linear or polynomial regression model to predict new values would be the best suited to this data, instead of a classification model.

Statistical Question

How can we predict the number of bike rentals on a given day, given environmental and seasonal factors pertaining to that day?

Data Characteristics

The following are the variables included in the bike sharing dataset:

Date: The date of each day being recorded.

Season: The season of each day being recorded. 1 corresponds to Spring, 2 corresponds to Summer, 3 corresponds to Autumn, and 4 corresponds to Winter.

Year: A value of '0' if the year was 2011, or '1' if the year was 2012.

Month: A value from 1 to 12 corresponding to which month of the year it was.

Holiday: A value of '1' if the day is a holiday, or '0' if the day is not a holiday.

Weekday: A value from 0 to 6 corresponding to which day of the week it was. 0 Corresponds to Sunday, 1 to Monday, et cetera.

Working Day: A value of '1' if the day is neither a weekend nor a holiday, or '0' if the day was a weekend or a holiday.

Weather: The weather of each day being recorded. 1 corresponds to Clear, Few Clouds, or Partly Cloudy. 2 Corresponds to Mist + Cloudy, Mist + Broken Clouds, Mist + Few Clouds, or Mist. 3 corresponds to Light Snow, Light Rain + Thunderstorm + Scattered CLouds, or Light Rain + Scattered Clouds. 4 corresponds to Heavy Rain + Ice Pellets + Thunderstorm + Mist, or Snow + Fog. However, I noticed when viewing the data that there are no instances of a category '4' day for this variable, and thus it only has values of 1, 2, or 3.

Temperature: The temperature, normalized, in Celsius.

Temperature (Feels Like): The 'Feeling Temperature,' normalized, in Celsius.

Humidity: Humidity, normalized.

Wind speed: Wind speed, normalized.

Casual Users: Count of bike rentals from 'casual' users, who are not registered with the rental company.

Registered Users: Count of bike rentals from users who are registered with the rental company.

User Count: Count of total bike rentals, including both casual users and registered users.

Initial Exploration

To begin with, I made plots of each individual explanatory variable in order to see if any patterns were immediately apparent. I chose not to examine the first two columns, "instant" and "date," because they both function as indices for this data and thus would not be useful for a regression model. I made boxplots for the categorical variables: season, year, month, holiday, weekday, working day, and weather. I made scatterplots for the continuous variables: temperature, temperature (feels like), humidity, and wind speed. I colored the points on these scatterplots first by season, and then by weather, because from the boxplots these appeared to be the most relevant to explaining changes in the rental count. For the season colors, I chose to use yellow for spring, green for summer, red for autumn, and blue for winter. For the weather colors, I chose to use green for 1, a clear or partly cloudy day, blue for 2, a misty or cloudy day, and red for 3, a rainy or snowy day. I then made separate scatterplots for the continuous variables that showed the relationships between these explanatory variables and both registered and casual users. For these, registered users were colored in blue and casual users were colored in red.

These initial graphs show that the season, the weather, and the temperature appear to have the most significant effect on the number of bike rentals on a given day. The graphs also indicate that there is a significant difference in the number of registered users versus the number of casual users. Because the total rental count for each day is simply the sum of the number of registered users and the number of casual users, we cannot use these variables to predict total count. So, it appears that we may predict the number of bike rentals far more accurately if we create two models from the data: one for registered users, and one for casual users.

Creating a Linear Model

To begin with, I made a simple linear model that uses each of the 11 predictor variables to predict the total rental count. These predictor variables are season, year, month, holiday, weekday, working day, weather, temperature, temperature (feels like), humidity, and wind speed. I also made two separate simple linear models that use these predictor variables to predict the rental count of casual users and the rental count of registered users. I then calculated the variance inflation factor of each variable in these models to check for multicollinearity before proceeding.

Intuitively, I felt that it would be necessary to remove some variables in order to create a better model. For example, the "month" variable and the "season" variable are basically redundant; knowing what month it is tells us what season it is. I also felt that the holiday variable should probably be removed, because knowing the day of the week and whether or not it was a working day gives us the same information.

The variance inflation factors show that the temperature variable and the temperature (feels like) variable have an extremely high collinearity. This makes sense, because these variables are almost always going to be quite close to each other. I therefore chose to remove the 'temperature' variable from each model and only use 'temperature (feels like),' as the latter appeared to be more statistically significant in each model.

I then chose to remove the 'month' and 'holiday' variables from each model, because they appeared to not be statistically significant. As I discussed earlier, it makes sense that these variables are superfluous given the other data available for the model. I also removed the 'season' variable from the model that only predicted the number of rentals from casual users, because it proved to not be statistically significant.

The resulting linear model to predict the total count of bike rentals then uses the following variables: season, year, weekday, working day, weather, temperature (feels like), humidity, and windspeed. It has a multiple r-squared value of 0.7963, which is good. However, the residual standard error is 879.2 on 722 degrees of freedom, which is quite high. We would hope to produce a model with lower error than this.

The linear model to predict the count of bike rentals from registered users uses the following variables: season, year, weekday, working day, weather, temperature (feels like), humidity, and windspeed. It has a multiple r-squared value of 0.8147, which is good.The residual standard error is 675.4 on 722 degrees of freedom. Like the previous model, this error is quite high, although it is better than the previous model.

The linear model to predict the count of bike rentals from casual users uses the following variables: year, weekday, working day, weather, temperature (feels like), humidity, and windspeed. It has a multiple r-squared value of 0.6797, which is not quite as good as the previous two models. The residual standard error is 390.5 on 723 degrees of freedom. This is better than the previous two models, but still quite high.

The extremely high variability in the residuals of these models indicates the need for a better model.

Creating a natural spline model

In order to better predict the number of bike rentals, I then chose to create two models which use natural splines. I chose to discard the idea of predicting the total count and only create two models, one for the rental count from registered users and one for the rental count from casual users. I felt that because total count is simply the sum of registered and casual renters, it was superfluous to attempt to create that model as well as the other two. From the graphs that I created, it seems apparent that there is a significant difference between casual and registered users, and thus it seems appropriate to model these two separately.

The natural spline models use the same variables as the original linear models. The model for registered users uses season, year, weekday, working day, weather, temperature (feels like), humidity, and windspeed. The model for casual users uses year, weekday, weather, temperature (feels like), humidity, and windspeed. Because some of the data was categorical, I did not need to use splines for some variables. I chose to use splines on the continuous data: temperature, humidity, and windspeed. I chose four knots for each variable.

The natural spline model to predict the count of bike rentals from registered users has a multiple r-squared value of 0.8733. The residual standard error is 561.3 on 713 degrees of freedom. This model performs significantly better than the linear model previously created. The error is much lower, and the coefficient of determination is higher. From these results we can estimate that about 87% of the variance in the number of bike rentals from registered users can be explained by this model.

The natural spline model to predict the count of bike rentals from casual users has a multiple r-squared value of 0.749. The residual standard error is 347.8 on 714 degrees of freedom. This model performs significantly better than the linear model previously created.The error is lower and the coefficient of determination is higher. From these results we can estimate that about 74.9% of the variance in the number of bike rentals from casual users can be explained by this model.

Conclusion

In this project, I created several different statistical models to predict the number of bike rentals on any given day given a set of environmental and seasonal variables pertaining to that day. I created a generic linear model, and then performed significance testing to reduce the number of dimensions needed for the model. I then split the model into two to predict the number of rentals from casual users and the number of rentals from registered users separately. This was done to produce a more accurate number of the total number of bike rentals.

These linear models had a high amount of error, so I created models using natural splines to predict the number of rentals from casual and registered users separately. These two models performed significantly better than the original linear model.

Possible next steps for this project could include testing different numbers of splines, or perhaps testing a different type of predictive model, such as a generalized additive model.
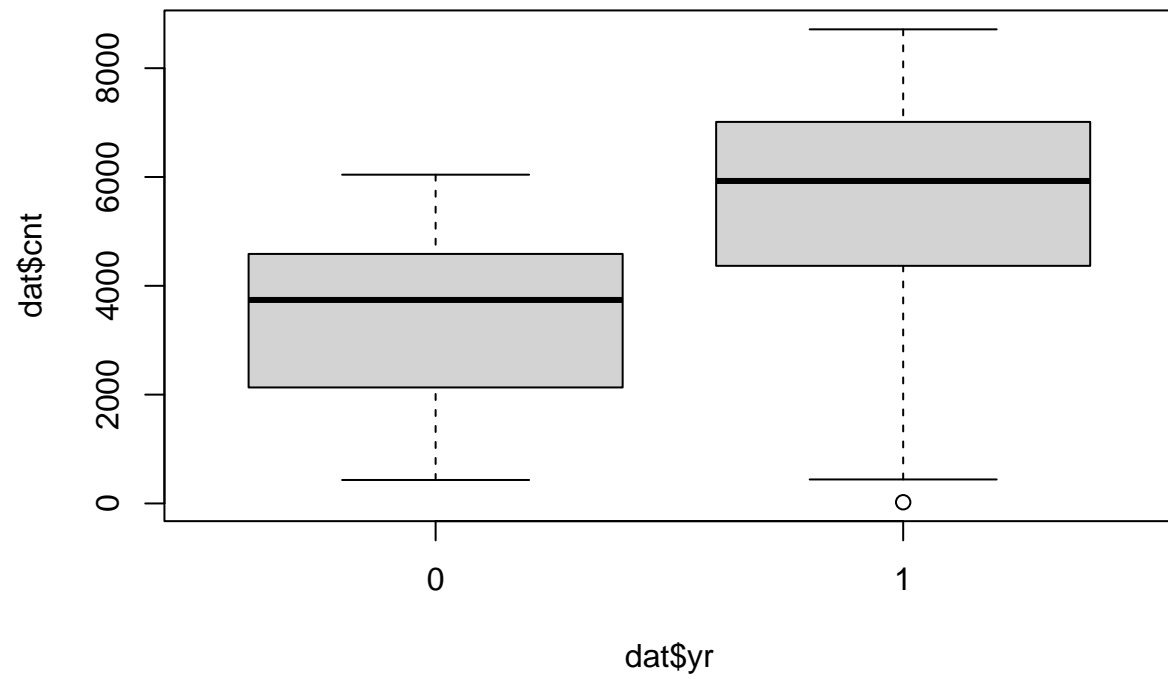
```r
dat <- read.csv("./day.csv")

# Initial boxplots for categorical explanatory variables
boxplot(dat$cnt ~ dat$season)
```
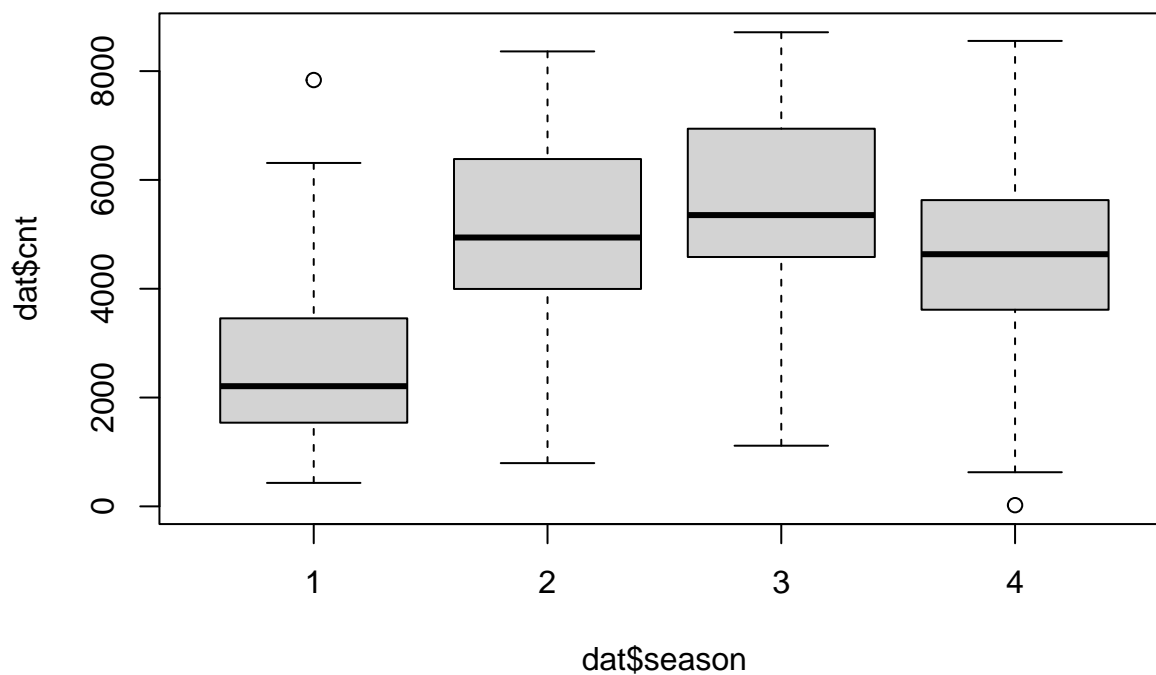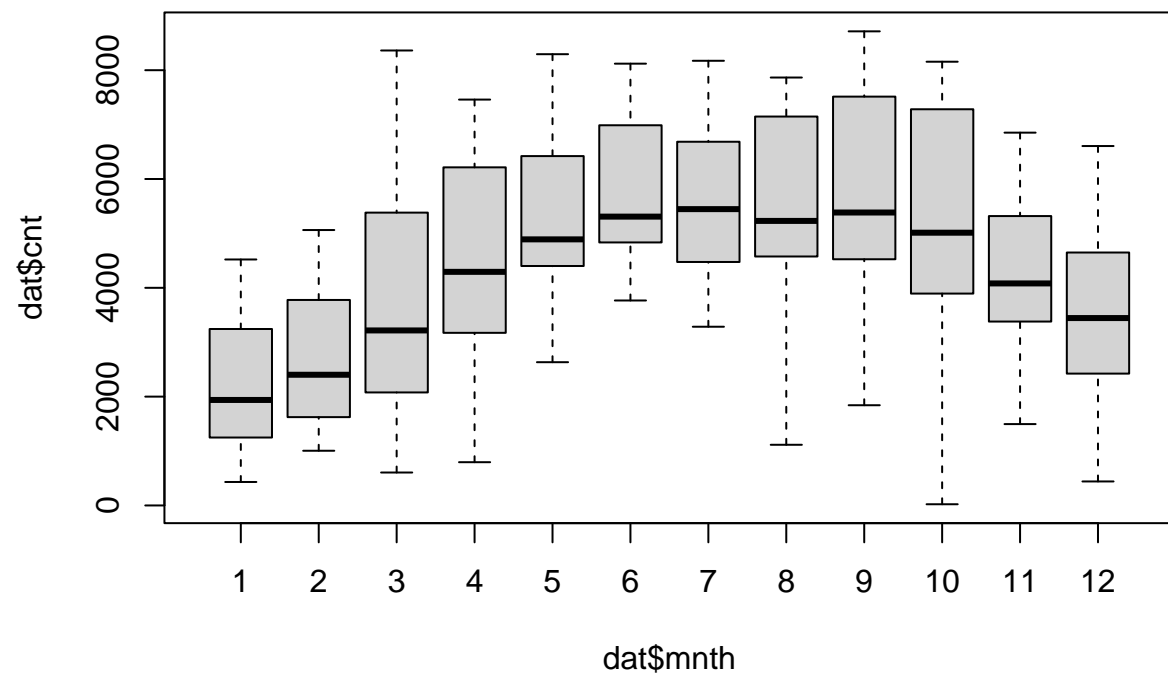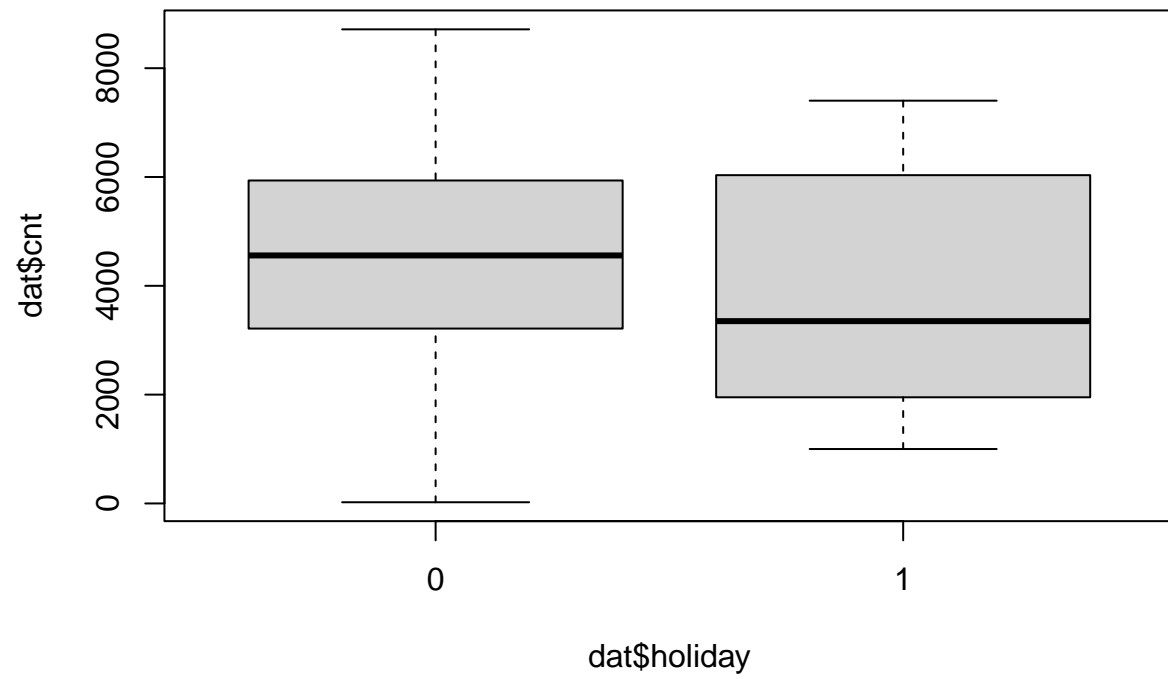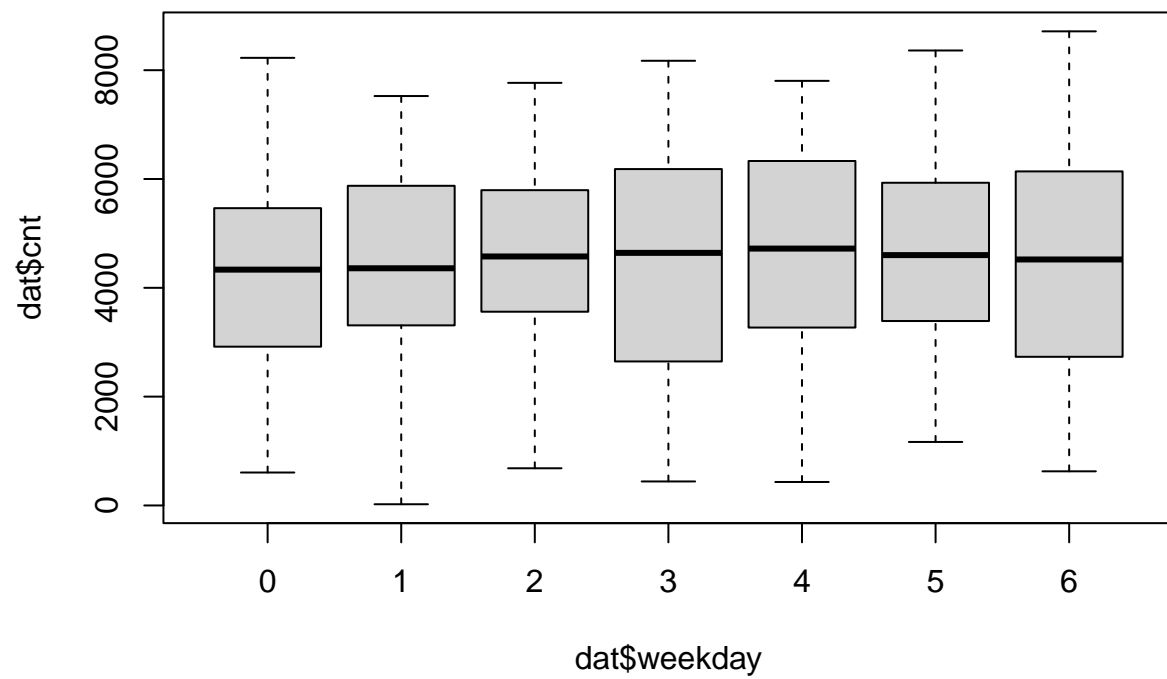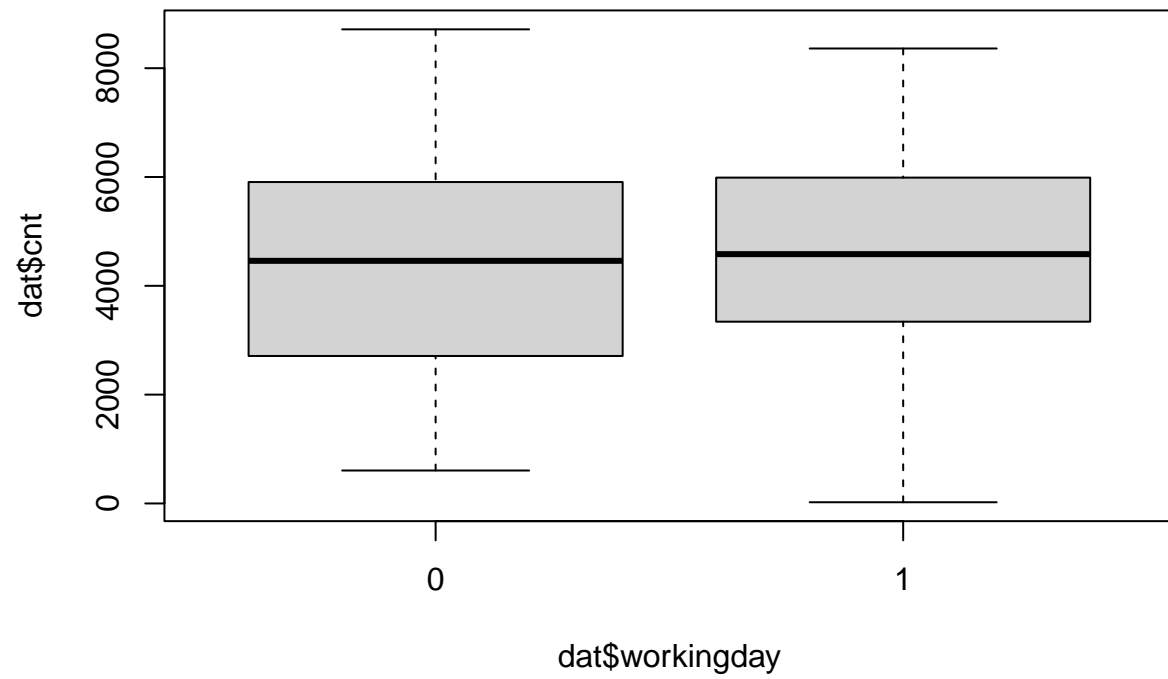


```r
boxplot(dat$cnt ~ dat$yr)
```

```
boxplot(dat$cnt ~ dat$mnth)
```

```
boxplot(dat$cnt ~ dat$holiday)
```

```
boxplot(dat$cnt ~ dat$weekday)
```

```
boxplot(dat$cnt ~ dat$workingday)
```

```
boxplot(dat$cnt ~ dat$weathersit)
```
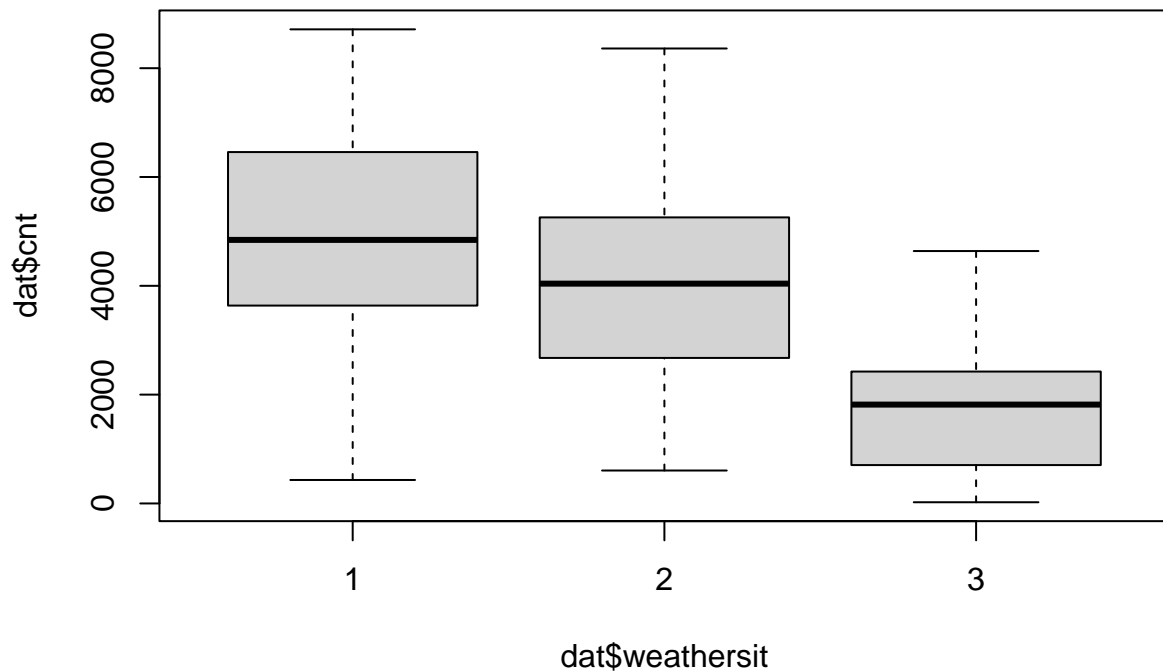
```r
# Creating a new column that colors the data by which season it was
dat$seasoncolor <- "yellow"
dat$weathercolor <- "green"
for(i in 1:731){
  if (dat[i, 3] == 2){
    dat[i,'seasoncolor'] = "green"
  } else if (dat[i, 3] == 3){
    dat[i,'seasoncolor'] = "red"
  } else if (dat[i,3] == 4){
    dat[i, 'seasoncolor'] = "blue"
  }

  if (dat[i, 9] == 2){
    dat[i,'weathercolor'] = "blue"
  } else if (dat[i, 9] == 3){
    dat[i,'weathercolor'] = "red"
  }
}

# Scatterplots of continuous data, colored by season
plot(dat$temp, dat$cnt, col=dat$seasoncolor)
```
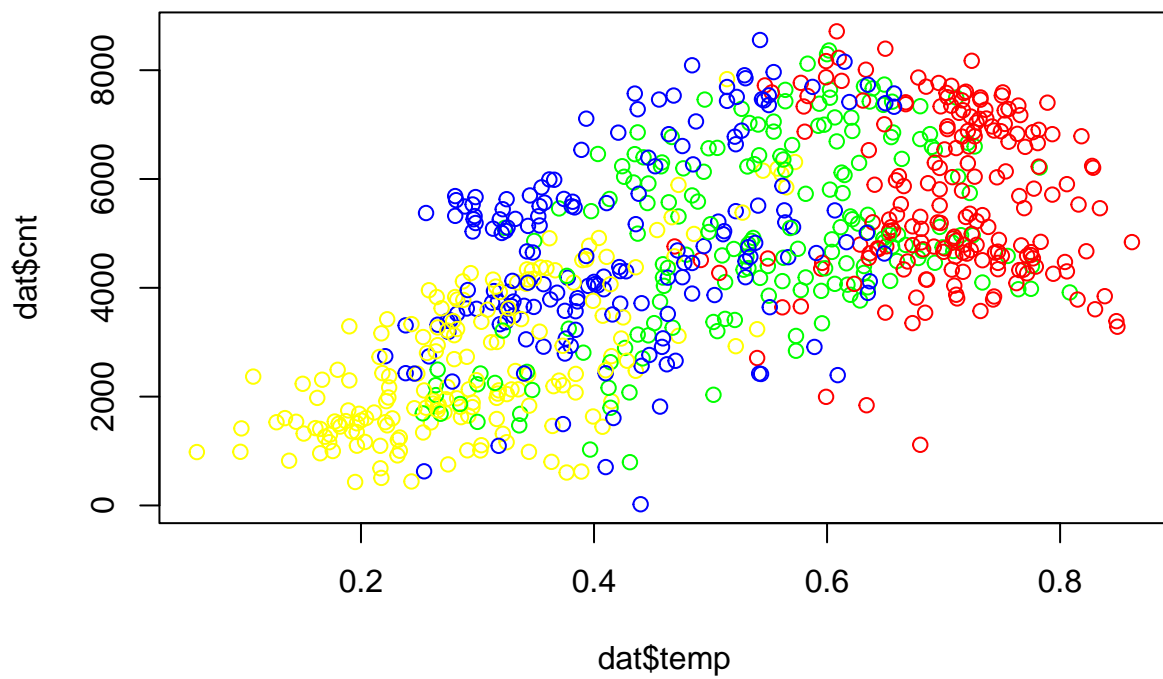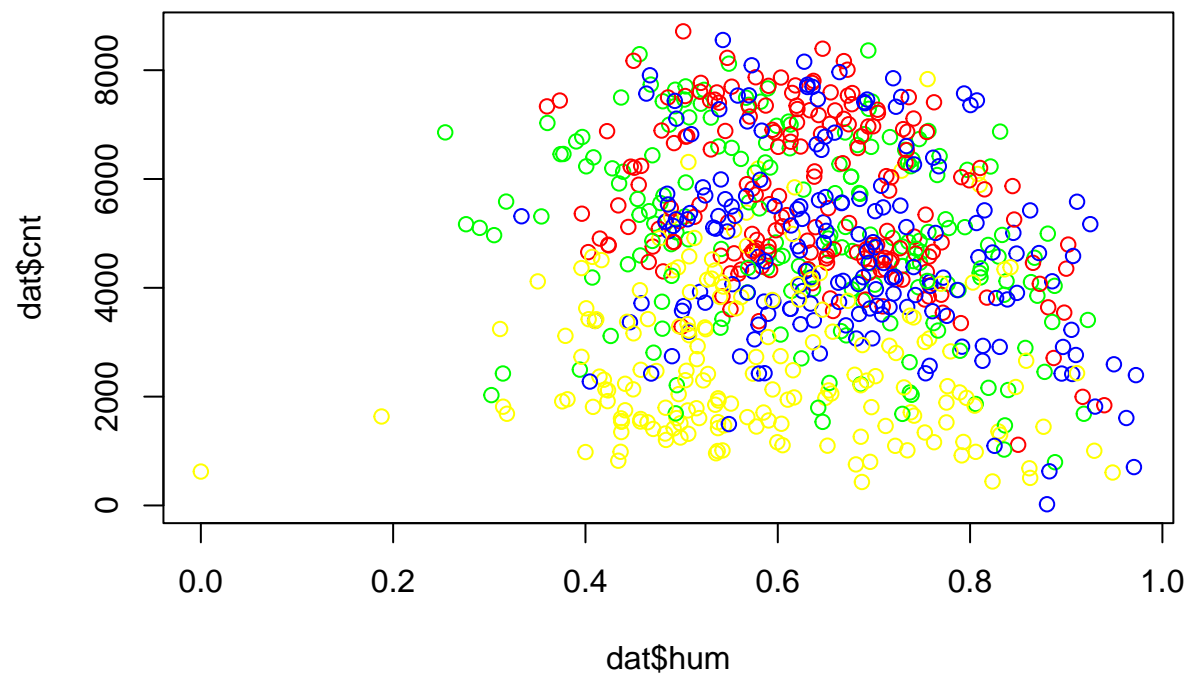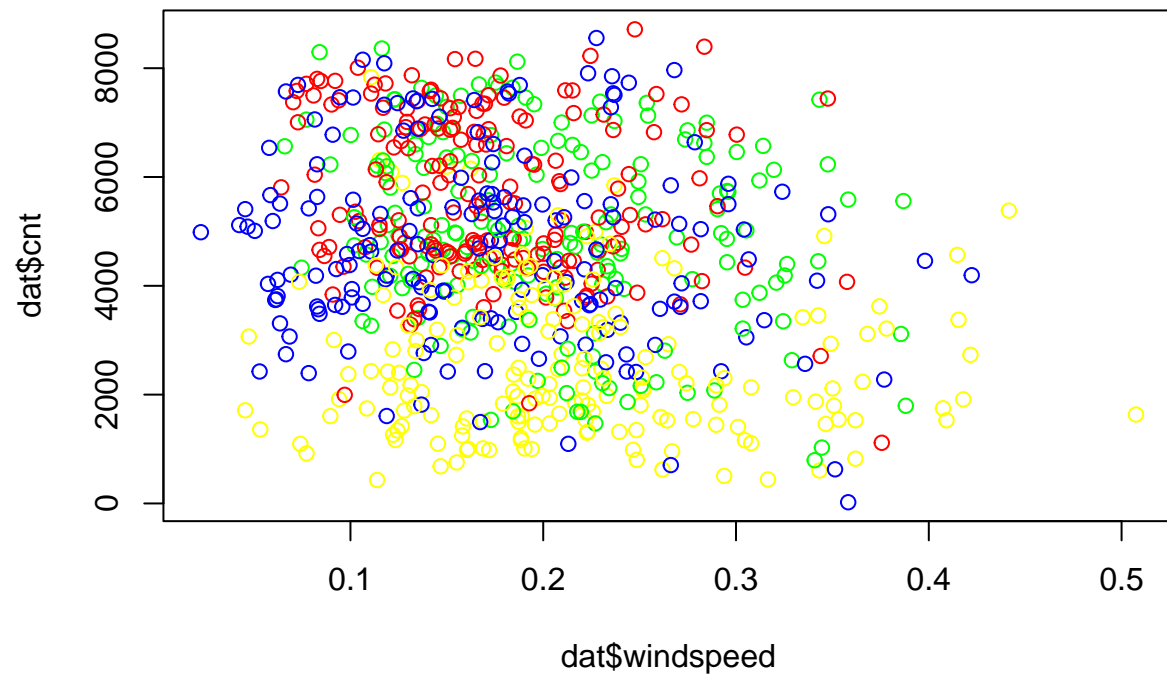
```
plot(dat$atemp, dat$cnt, col=dat$seasoncolor)
```
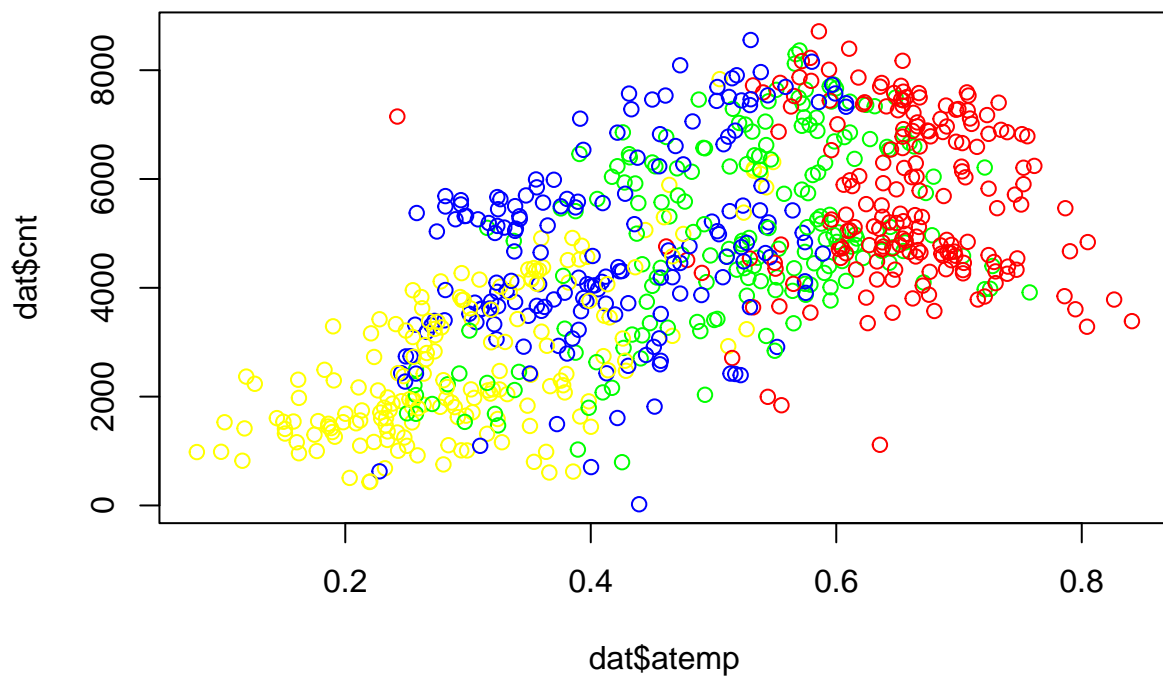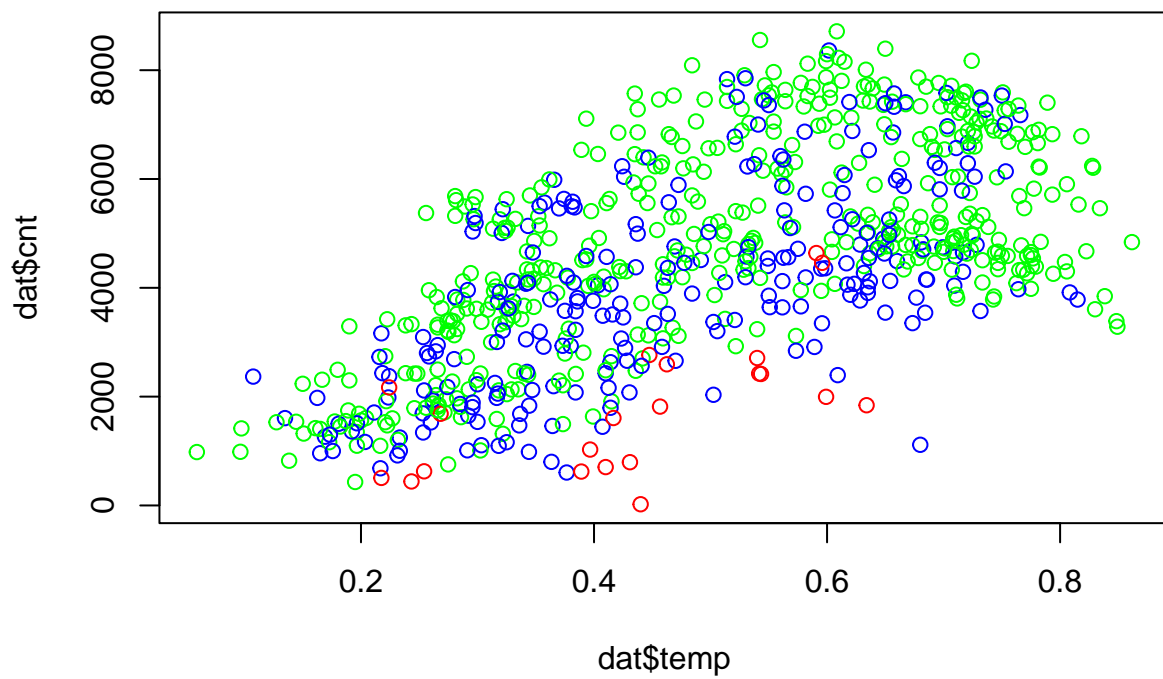
```
plot(dat$hum, dat$cnt, col=dat$seasoncolor)
```
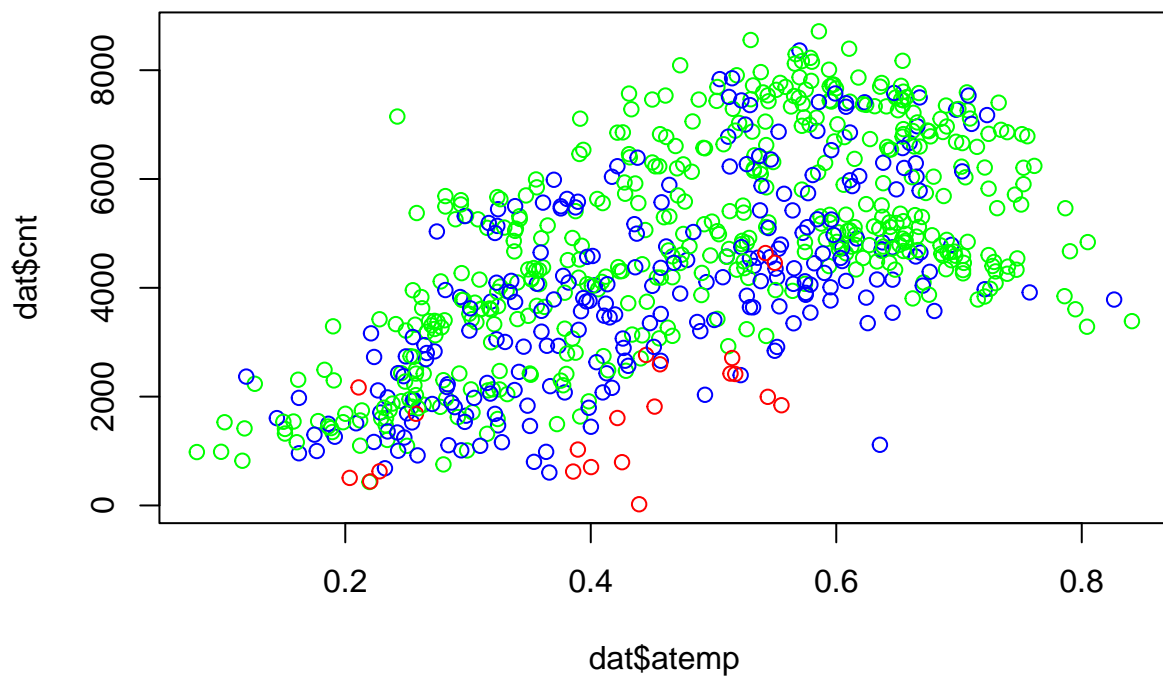
```
plot(dat$windspeed, dat$cnt, col=dat$seasoncolor)
```

```
# Scatterplots of continuous data, colored by weather
plot(dat$temp, dat$cnt, col=dat$weathercolor)
```

```
plot(dat$atemp, dat$cnt, col=dat$weathercolor)
```

```
plot(dat$hum, dat$cnt, col=dat$weathercolor)
```

```
plot(dat$windspeed, dat$cnt, col=dat$weathercolor)
```
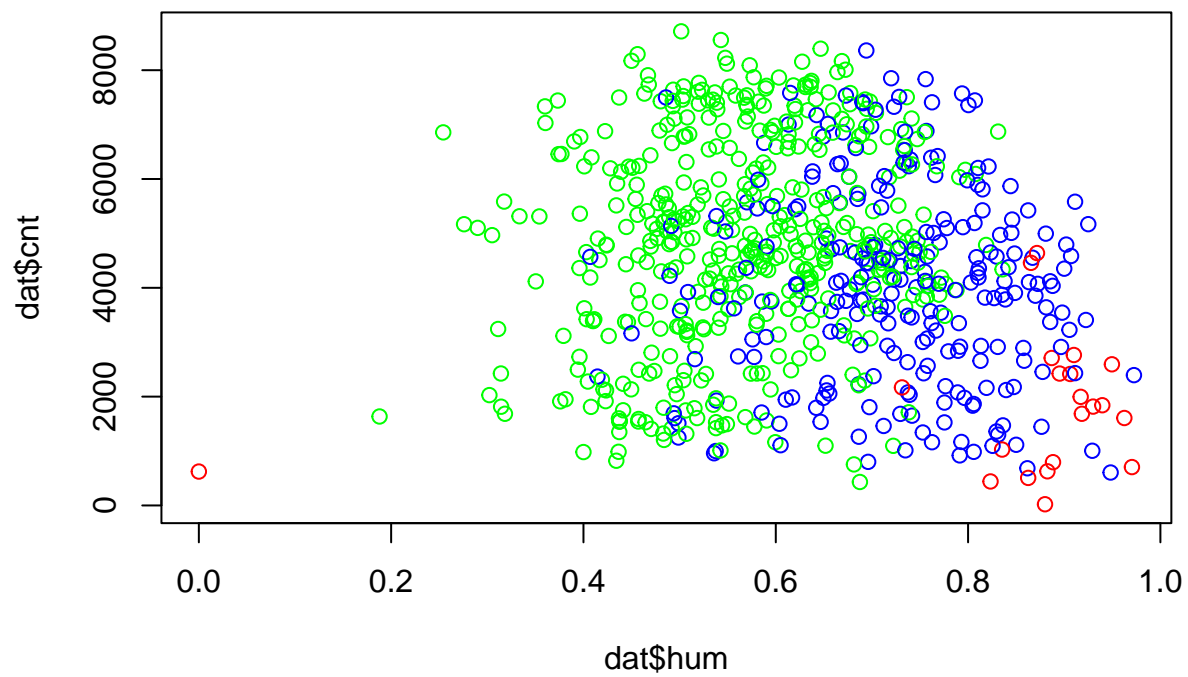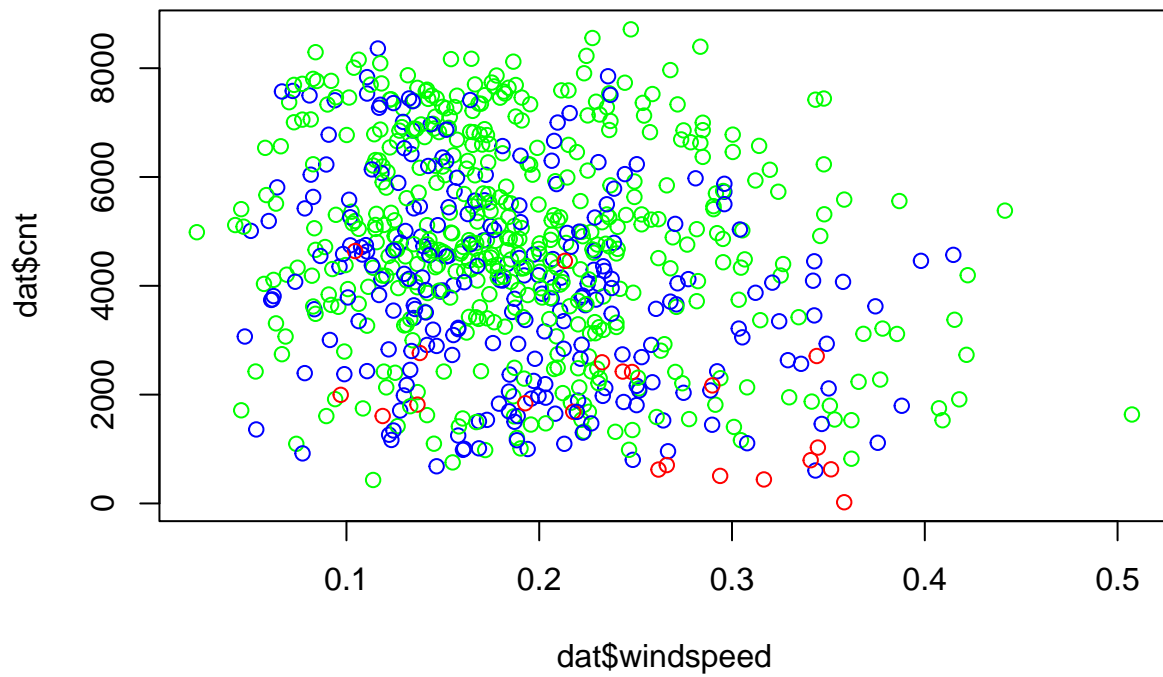
```
# Scatterplots of continuous data, colored by casual and registered users
plot(dat$temp, dat$cnt, col="gray")
points(dat$temp, dat$casual, col="red")
points(dat$temp, dat$registered, col="blue")
```

```
plot(dat$atemp, dat$cnt, col="gray")
points(dat$atemp, dat$casual, col="red")
points(dat$atemp, dat$registered, col="blue")
```
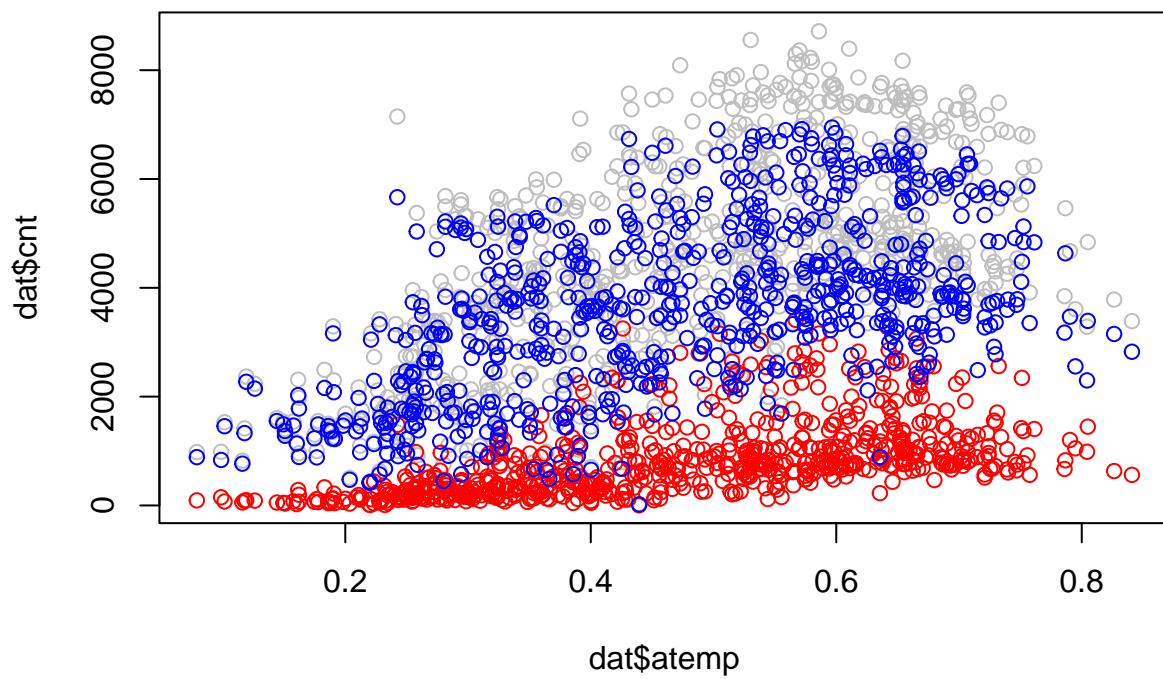
```
plot(dat$hum, dat$cnt, col="gray")
points(dat$hum, dat$casual, col="red")
points(dat$hum, dat$registered, col="blue")
```
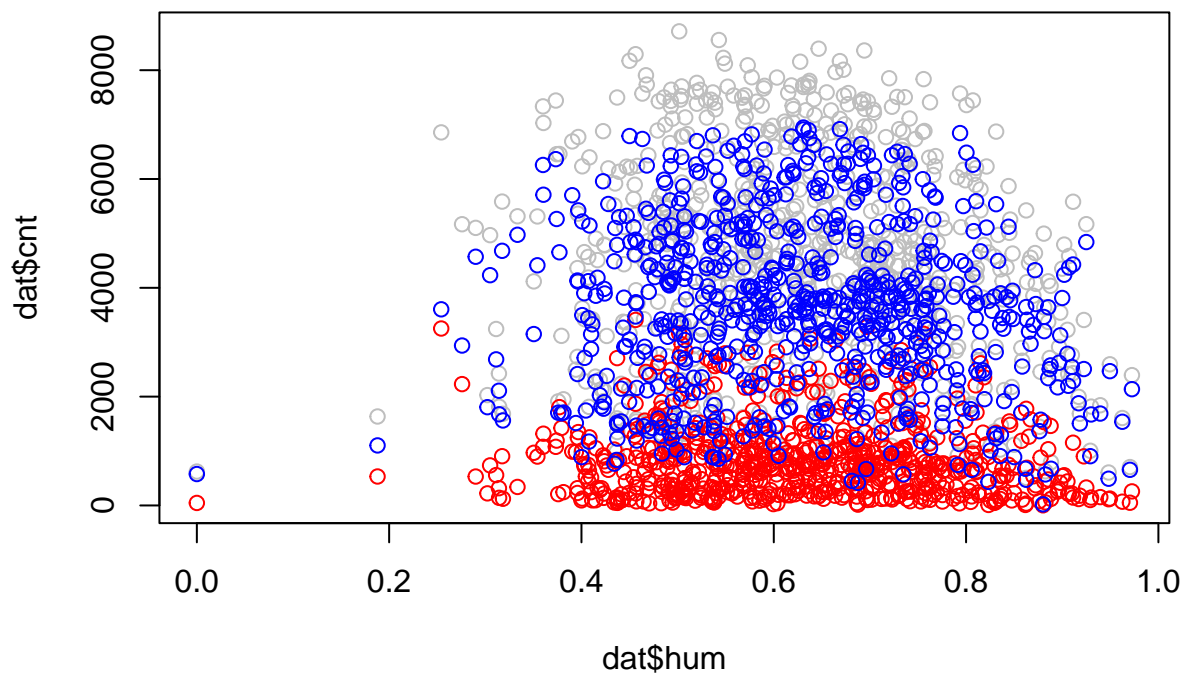
```
plot(dat$windspeed, dat$cnt, col="gray")
points(dat$windspeed, dat$casual, col="red")
points(dat$windspeed, dat$registered, col="blue")
```
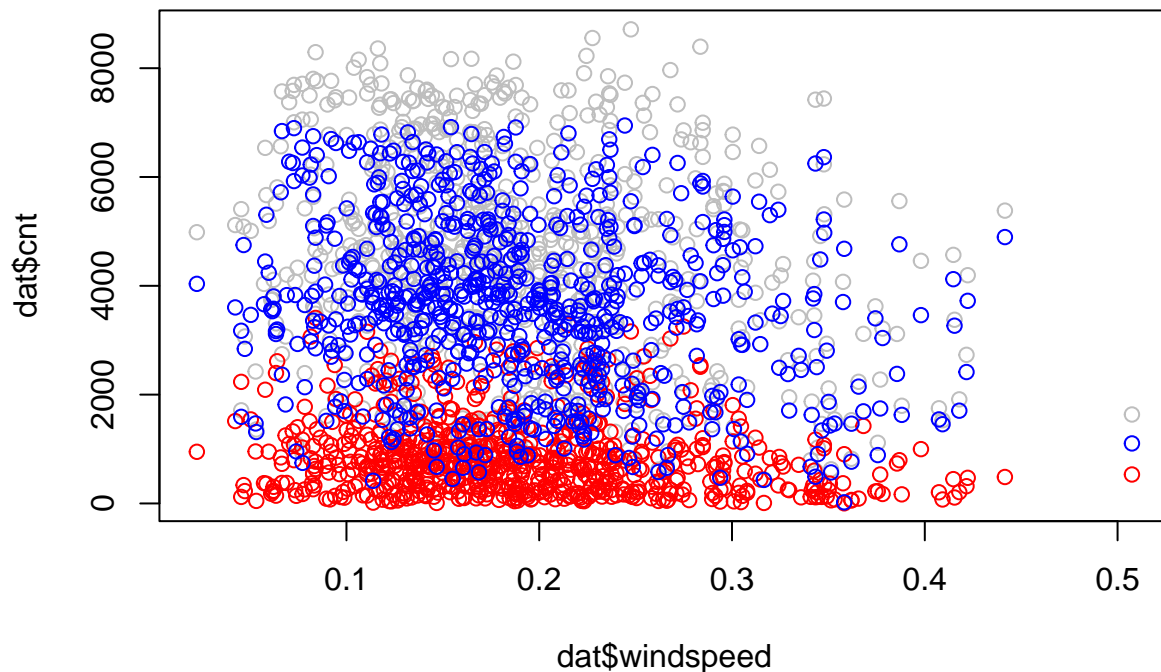
```
library(car)
```

```
## Loading required package: carData
# Initial linear models using every predictor variable, and variance inflation factors.
total_lm1 <- lm(cnt ~ season + yr + mnth + holiday + weekday + workingday + weathersit + temp + atemp +
casual_lm1 <- lm(casual ~ season + yr + mnth + holiday + weekday + workingday + weathersit + temp + ate
registered_lm1 <- lm(registered ~ season + yr + mnth + holiday + weekday + workingday + weathersit + ter
summary(total_lm1)
```

```
##
## Call:
## lm(formula = cnt ~ season + yr + mnth + holiday + weekday + workingday +
##     weathersit + temp + atemp + hum + windspeed, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4143.6  -442.3    49.5   546.2  2946.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1469.00     240.22    6.115 1.58e-09 ***
## season        509.78      54.76    9.310  < 2e-16 ***
## yr           2040.70      65.19   31.306  < 2e-16 ***
## mnth          -38.98      17.08   -2.282  0.02276 *
## holiday      -518.99     201.04   -2.582  0.01003 *
## weekday        69.06      16.30    4.237 2.56e-05 ***
```

```
## workingday    120.36      72.01   1.671  0.09507 .
## weathersit   -610.99      78.36  -7.797 2.23e-14 ***
## temp         2028.92    1403.67   1.445  0.14877
## atemp        3573.27    1589.39   2.248  0.02487 *
## hum         -1018.86     314.00  -3.245  0.00123 **
## windspeed   -2557.57     456.28  -5.605 2.96e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 872.4 on 719 degrees of freedom
## Multiple R-squared:  0.8002, Adjusted R-squared:  0.7972
## F-statistic: 261.9 on 11 and 719 DF,  p-value: < 2.2e-16
```

```
vif(total_lm1)
```

```
##     season         yr       mnth    holiday    weekday workingday weathersit
##   3.548413   1.020253   3.333672   1.083126   1.024076   1.076392   1.748741
##       temp      atemp        hum  windspeed
##  63.321299  64.343361   1.918309   1.199259
```

```
summary(casual_lm1)
```

```
##
## Call:
## lm(formula = casual ~ season + yr + mnth + holiday + weekday +
##     workingday + weathersit + temp + atemp + hum + windspeed,
##     data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1258.79  -221.62   -13.06   179.80  1620.28
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  700.793    106.358   6.589 8.56e-11 ***
## season        61.824     24.244   2.550 0.010977 *
## yr           286.685     28.861   9.933  < 2e-16 ***
## mnth         -15.744      7.562  -2.082 0.037690 *
## holiday     -274.214     89.012  -3.081 0.002144 **
## weekday       26.364      7.216   3.653 0.000278 ***
## workingday  -828.251     31.882 -25.979  < 2e-16 ***
## weathersit  -113.049     34.696  -3.258 0.001174 **
## temp        1194.842    621.486   1.923 0.054931 .
## atemp        894.855    703.714   1.272 0.203920
## hum         -393.230    139.024  -2.829 0.004807 **
## windspeed   -862.054    202.020  -4.267 2.24e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 386.3 on 719 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6835
## F-statistic: 144.3 on 11 and 719 DF,  p-value: < 2.2e-16
```

```
vif(casual_lm1)
```

```
##     season         yr       mnth    holiday    weekday workingday weathersit
```

```
##    3.548413    1.020253    3.333672    1.083126    1.024076    1.076392    1.748741
##        temp       atemp         hum  windspeed
##   63.321299   64.343361    1.918309    1.199259
```

```
summary(registered_lm1)
```

```
##
## Call:
## lm(formula = registered ~ season + yr + mnth + holiday + weekday +
##     workingday + weathersit + temp + atemp + hum + windspeed,
##     data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3945.2  -350.7    77.2   430.5  1595.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   768.21     185.55   4.140 3.88e-05 ***
## season        447.95      42.30  10.591  < 2e-16 ***
## yr           1754.02      50.35  34.836  < 2e-16 ***
## mnth          -23.24      13.19  -1.761 0.078618 .
## holiday      -244.78     155.29  -1.576 0.115403
## weekday        42.70      12.59   3.392 0.000733 ***
## workingday    948.61      55.62  17.055  < 2e-16 ***
## weathersit   -497.94      60.53  -8.226 9.04e-16 ***
## temp          834.07    1084.23   0.769 0.441982
## atemp        2678.42    1227.69   2.182 0.029456 *
## hum          -625.63     242.54  -2.580 0.010091 *
## windspeed   -1695.52     352.44  -4.811 1.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 673.9 on 719 degrees of freedom
## Multiple R-squared:  0.8163, Adjusted R-squared:  0.8135
## F-statistic: 290.4 on 11 and 719 DF,  p-value: < 2.2e-16
```

```
vif(registered_lm1)
```

```
##      season          yr        mnth     holiday      weekday workingday weathersit
##    3.548413    1.020253    3.333672    1.083126    1.024076    1.076392    1.748741
##        temp       atemp         hum  windspeed
##   63.321299   64.343361    1.918309    1.199259
```

```
# Second set of linear models, using a reduced number of predictor variables
total_lm2 <- lm(cnt ~ season + yr + weekday + workingday + weathersit + atemp + hum + windspeed, data =
casual_lm2 <- lm(casual ~ yr + weekday + workingday + weathersit + atemp + hum + windspeed, data = dat)
registered_lm2 <- lm(registered ~ season + yr + weekday + workingday + weathersit + atemp + hum + windsp
summary(total_lm2)
```

```
##
## Call:
## lm(formula = cnt ~ season + yr + weekday + workingday + weathersit +
##     atemp + hum + windspeed, data = dat)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4164.7  -411.3    57.0   529.9  2950.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1319.21     232.56    5.673 2.03e-08 ***
## season        407.64      32.00   12.739  < 2e-16 ***
## yr           2037.50      65.67   31.024  < 2e-16 ***
## weekday        73.11      16.32    4.480 8.66e-06 ***
## workingday    170.93      70.29    2.432 0.015268 *
## weathersit   -598.01      78.87   -7.582 1.05e-13 ***
## atemp        5922.87     220.18   26.901  < 2e-16 ***
## hum         -1115.48     314.37   -3.548 0.000413 ***
## windspeed   -2433.01     453.59   -5.364 1.10e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 879.2 on 722 degrees of freedom
## Multiple R-squared:  0.7963, Adjusted R-squared:  0.794
## F-statistic: 352.8 on 8 and 722 DF,  p-value: < 2.2e-16
```

```
summary(casual_lm2)
```

```
##
## Call:
## lm(formula = casual ~ yr + weekday + workingday + weathersit +
##     atemp + hum + windspeed, data = dat)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -1259.88  -223.50    -3.96   177.89  1610.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  642.009    101.669    6.315 4.72e-10 ***
## yr           285.282     29.168    9.781  < 2e-16 ***
## weekday       28.824      7.247    3.978 7.66e-05 ***
## workingday  -802.367     31.217  -25.703  < 2e-16 ***
## weathersit  -107.784     35.021   -3.078  0.00216 **
## atemp       2307.452     93.527   24.671  < 2e-16 ***
## hum         -413.813    138.690   -2.984  0.00294 **
## windspeed   -832.448    199.479   -4.173 3.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 390.5 on 723 degrees of freedom
## Multiple R-squared:  0.6797, Adjusted R-squared:  0.6766
## F-statistic: 219.2 on 7 and 723 DF,  p-value: < 2.2e-16
```

```
summary(registered_lm2)
```

```
##
## Call:
## lm(formula = registered ~ season + yr + weekday + workingday +
##     weathersit + atemp + hum + windspeed, data = dat)
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3949.4  -341.3    83.5   426.9  1561.1
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   703.47     178.66   3.937 9.04e-05 ***
## season        387.13      24.58  15.748  < 2e-16 ***
## yr           1752.12      50.46  34.726  < 2e-16 ***
## weekday        44.40      12.54   3.541 0.000424 ***
## workingday    972.91      54.00  18.016  < 2e-16 ***
## weathersit   -491.28      60.59  -8.108 2.21e-15 ***
## atemp        3656.63     169.15  21.617  < 2e-16 ***
## hum          -678.43     241.52  -2.809 0.005104 **
## windspeed   -1641.18     348.48  -4.710 2.98e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 675.4 on 722 degrees of freedom
## Multiple R-squared:  0.8147, Adjusted R-squared:  0.8126
## F-statistic: 396.7 on 8 and 722 DF,  p-value: < 2.2e-16
```

```r
library(splines)
# Creating a natural spline model for registered users
reg.ns <- lm(registered ~ season + yr + weekday + workingday + weathersit + ns(atemp, 4) + ns(hum, 4) +
summary(reg.ns)
```

```
## 
## Call:
## lm(formula = registered ~ season + yr + weekday + workingday +
##     weathersit + ns(atemp, 4) + ns(hum, 4) + ns(windspeed, 4),
##     data = dat)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3663.3  -272.0    61.4   346.1  1391.5
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -232.13     512.66  -0.453   0.6508
## season             365.91      21.57  16.962  < 2e-16 ***
## yr                1691.67      42.82  39.504  < 2e-16 ***
## weekday             47.63      10.48   4.544 6.49e-06 ***
## workingday         935.36      45.24  20.677  < 2e-16 ***
## weathersit        -371.77      54.06  -6.877 1.34e-11 ***
## ns(atemp, 4)1     1876.26     170.60  10.998  < 2e-16 ***
## ns(atemp, 4)2     2539.71     146.02  17.392  < 2e-16 ***
## ns(atemp, 4)3     1628.75     398.14   4.091 4.79e-05 ***
## ns(atemp, 4)4      228.94     195.66   1.170   0.2424
## ns(hum, 4)1        666.55     412.54   1.616   0.1066
## ns(hum, 4)2         61.19     270.90   0.226   0.8214
## ns(hum, 4)3        645.12     882.37   0.731   0.4649
## ns(hum, 4)4       -994.62     234.77  -4.237 2.57e-05 ***
## ns(windspeed, 4)1 -158.72     171.24  -0.927   0.3543
```

```
## ns(windspeed, 4)2   -146.25      160.40  -0.912    0.3622
## ns(windspeed, 4)3   -758.18      408.96  -1.854    0.0642 .
## ns(windspeed, 4)4  -1391.73      289.27  -4.811 1.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 561.9 on 713 degrees of freedom
## Multiple R-squared:  0.8733, Adjusted R-squared:  0.8703
## F-statistic: 289.1 on 17 and 713 DF,  p-value: < 2.2e-16
```

```r
# Creating a natural spline model for casual users
cas.ns <- lm(casual ~ yr + weekday + workingday + weathersit + ns(atemp, 4) + ns(hum, 4) + ns(windspeed
summary(cas.ns)
```

```
##
## Call:
## lm(formula = casual ~ yr + weekday + workingday + weathersit +
##     ns(atemp, 4) + ns(hum, 4) + ns(windspeed, 4), data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1020.24  -220.58   -26.86   187.92  1410.90
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        983.375    317.164   3.101  0.00201 **
## yr                 271.198     26.503  10.233  < 2e-16 ***
## weekday             31.725      6.488   4.890 1.25e-06 ***
## workingday        -820.221     27.994 -29.300  < 2e-16 ***
## weathersit         -83.944     33.462  -2.509  0.01234 *
## ns(atemp, 4)1     1062.673    100.184  10.607  < 2e-16 ***
## ns(atemp, 4)2     1391.395     88.370  15.745  < 2e-16 ***
## ns(atemp, 4)3     1092.243    239.199   4.566 5.85e-06 ***
## ns(atemp, 4)4      548.008    118.409   4.628 4.38e-06 ***
## ns(hum, 4)1       -315.165    254.680  -1.237  0.21631
## ns(hum, 4)2       -356.216    167.576  -2.126  0.03387 *
## ns(hum, 4)3       -876.766    545.603  -1.607  0.10850
## ns(hum, 4)4       -657.172    144.093  -4.561 6.00e-06 ***
## ns(windspeed, 4)1 -215.176    104.178  -2.065  0.03924 *
## ns(windspeed, 4)2   57.323     98.804   0.580  0.56198
## ns(windspeed, 4)3 -659.008    249.587  -2.640  0.00846 **
## ns(windspeed, 4)4 -1118.366   178.064  -6.281 5.85e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 347.8 on 714 degrees of freedom
## Multiple R-squared:  0.749,  Adjusted R-squared:  0.7434
## F-statistic: 133.2 on 16 and 714 DF,  p-value: < 2.2e-16
```